

스마트폰 기반 실시간 학습 행동 인식 및 자동 기록 시스템

유 민 경¹ · 이 하 늘¹ · 오 지 연¹ · 박 유 정¹ · 양 규 리¹ · 이 경 미^{2*}¹덕성여자대학교 컴퓨터공학전공 학사과정²덕성여자대학교 컴퓨터공학전공 교수

Smartphone-based Real-Time Study Behavior Recognition and Automatic Study Logging System

Min-Kyeong Yu¹ · Ha-Neul Lee¹ · Ji-Yeon Oh¹ · Yoo-Jung Park¹ · Gyu-Ri Yang¹ · Kyoung-Mi Lee^{2*}¹Bachelor's Course, Department of Computer Engineering, Duksung Women's University, Seoul 01369, Korea²Professor, Department of Computer Engineering, Duksung Women's University, Seoul 01369, Korea

[요 약]

본 논문은 스마트폰 카메라 기반으로 손 모양과 상체 자세를 분석하여 학습자의 실제 학습 행동을 자동 인식하는 모델을 제안한다. 이를 위해 MediaPipe로부터 손·상체·얼굴 랜드마크를 추출하고, 펜 킴 손 분류기와 여섯 가지 상체 자세를 판별하는 자세 분류기를 각각 MLP 구조로 개발하였으며, 특징 구성(절대좌표, 상대좌표, 관절각, 얼굴 특징 등)과 네트워크 깊이에 따른 성능 변화를 체계적으로 비교하였다. 실험 결과, 상대좌표·관절각·심층 구조를 결합한 손 인식 모델은 정확도 93.7%를 달성하였고, 상체 자세와 FaceMesh 정보를 통합한 모델은 84%의 정확도를 보였다. 두 모델의 출력을 규칙 기반으로 통합함으로써 네 가지 학습 행동(집중, 집중력 저하, 수면, 공부 중단)을 86.2% 정확도로 분류할 수 있었다. 이를 모바일 앱에 적용하여 실시간 기록·시각화·피드백 기능을 구현하여, 스마트폰만으로도 추가 장비 없이 학습 행동을 정량적으로 모니터링할 수 있음을 보여준다.

[Abstract]

This paper presents a smartphone-based system that automatically recognizes study behaviors by analyzing the hand shape and upper-body posture. The system extracted hand, pose, and face landmarks using MediaPipe and developed two MLP classifiers: a pen-in-hand detector and posture classifier distinguishing six study-related postures. We compared multiple feature configurations—including absolute and relative coordinates, joint angles, and facial landmarks—and evaluated changes in the network depth. Experiments showed that the hand model combining relative coordinates and joint angles achieved 93.7% accuracy, and the optimized posture model integrating upper-body pose and FaceMesh features achieved 84% accuracy. A rule-based fusion of the two models classified four study behavior states—focus, reduced focus, sleep, and stop studying—with 86.2% accuracy. The system was deployed in a mobile application supporting real-time logging, visualization, and feedback, demonstrating that an accurate study behavior monitoring was feasible using only a smartphone camera.

색인어 : 행동 인식, 손 모양 분류, 자세 분류, 스마트폰 모니터링, 미디어파이프**Keyword** : Behavior Recognition, Hand-Shape Classification, Posture Classification, Smartphone Monitoring, Mediapipe<http://dx.doi.org/10.9728/dcs.2026.27.2.313>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 15 December 2025; Revised 23 January 2026

Accepted 09 February 2026

*Corresponding Author: Kyoung-Mi Lee

Tel: +82-2-901-8348

E-mail: kmlee@duksung.ac.kr

I. 서론

대한민국은 세계적으로 높은 교육 참여율을 보이며, 2024년 OECD 기준 25~34세 고등교육 이수율 1위를 기록한다 [1]. 이러한 교육 환경 속에서 학습 효율성 향상을 지원하는 학습 관리 도구에 대한 수요는 꾸준히 증가하고 있다. 그러나 기존 타이머 기반 학습 관리 애플리케이션은 사용자의 수동 입력에 의존하므로, 실제 학습 여부나 집중 상태를 반영하지 못한다는 구조적 한계를 가진다.

이러한 한계를 보완하기 위해 시선 추적, 표정 분석 등 영상 기반 학습 상태 인식 연구가 활발히 수행되었다. 시선·표정 기반 모델은 정면 촬영 조건에서는 일정 수준의 성능을 보이나, 필기·상체 움직임이 잦고 시선이 분산되는 실제 학습 환경에서는 적용이 제한적이다. 더불어 얼굴·시선 정보만으로는 학습 여부, 집중도, 졸음과 같은 핵심 행동 지표를 직접적으로 판단하기 어려워 해석상의 제약이 존재한다.

학습 중 나타나는 행동은 필기 여부와 같은 학습 행위(action)와 앉은 상태에서 나타나는 상체 형태의 학습 자세(posture)가 결합되어 나타나며, 두 요소를 함께 고려할 때 학습 지속성, 집중도 및 피로도 변화를 보다 정밀하게 해석할 수 있다. 그럼에도 불구하고 기존 연구는 단일 신호 기반 접근에 주로 의존하고 있어, 학습 행위와 자세 정보를 통합적으로 활용한 학습 행동(behavior) 인식 연구는 제한적인 수준에 머물러 있다. Human Activity Recognition(HAR) 분야 역시 보행이나 제스처와 같은 비교적 큰 신체 동작을 중심으로 발전해 왔으며, 장시간 착석 환경에서 발생하는 미세한 상체 변화나 지속적인 자세 편차는 기존 HAR 모델에서 충분히 표현되지 않는 것으로 보고되었다[2]. 이러한 한계는 학습 환경에서 요구되는 정밀한 행동 분류의 어려움을 의미하며, 이를 보완하기 위한 방안으로 최근에는 서로 다른 신체 신호를 결합하여 세밀한 행동 차이를 인식하려는 멀티모달 행동 인식 기법이 주목받고 있다[3].

본 논문은 스마트폰 카메라만을 이용하여 손 모양과 상체 자세를 동시에 분석하고 이를 통합해 학습 행동을 네 가지 상태—‘집중(Focus)’, ‘집중력 저하(Reduced Focus)’, ‘수면(Sleep)’, ‘공부 중단(Stop Studying)’—로 자동 분류하는 새로운 학습 행동 인식 모델을 제안한다. 제안된 모델은 MediaPipe 기반 랜드마크 추출과 다층 퍼셉트론(MLP) 기반 분류 구조를 활용해 모바일 환경에서 실시간 처리가 가능하도록 설계된다. 이를 기반으로 자동 학습 기록, 시각화, AI 기반 피드백 기능을 제공하는 모바일 시스템을 구현함으로써 기존 타이머 중심 학습 관리 방식의 한계를 극복하고, 스마트폰만으로 학습 행동의 정량적 분석이 가능함을 보여준다.

본 논문의 구성은 다음과 같다. II장에서는 관련 연구를 검토하고, III장에서는 손·자세 인식 모델과 통합 알고리즘을 설명한다. IV장에서는 구현된 모바일 시스템을 기술하며, V장에서 결론과 향후 연구 방향을 제시한다.

II. 관련 연구

학습 행동 인식 연구는 학습자의 주의(집중), 참여도, 졸음 등을 파악하기 위해 얼굴·시선, 자세, 생리신호 등 다양한 신체적 지표를 활용해 발전해 왔다. 그러나 기존 접근들은 공통적으로 실제 학습 행위(예: 필기)의 존재 여부를 직접 판단하지 못한다는 구조적 한계를 가진다. 이 점은 학습자의 실제 공부 여부와 학습 태도를 동시에 고려하는 통합적 접근의 필요성을 뒷받침한다.

첫째, 웹캠 기반 얼굴·시선 인식 연구는 눈 깜빡임, 시선 방향, 얼굴 움직임 등을 분석해 집중 상태를 추정한다. Webcam-based Attention Tracking[4], DeepFace-Attention[5], 그리고 Hu & Gao[6]는 표정·시각 특징을 활용해 집중도·참여도를 평가하였다. 그러나 이러한 방식은 정면 얼굴이 안정적으로 촬영되는 온라인 환경에 최적화되어 있으며, 오프라인 학습처럼 시선이 교재·노트·모니터 등으로 분산되고 얼굴이 부분적으로 가려지는 상황에서는 정확도가 크게 저하된다. 일부 측면 카메라 추가 연구가 보고되었으나 [4], 다중 카메라 설치가 필요해 실제 학습 공간에서는 활용성이 낮다.

둘째, 자세 기반 인식 연구는 상체 기울기, 어깨 정렬, 목 방향 등을 분석해 집중·피로도를 추정한다. Zhang et al.[7]과 Ji et al.[8]은 HRNet 및 경량 포즈 인식 네트워크를 활용해 교실 환경에서 학생 자세를 고정밀로 분류하였으며, Dey et al.[9]은 어텐션 기반 네트워크로 온라인 학습 행동 인식 성능을 향상시켰다. 이러한 연구들은 다수의 자세·행동 클래스를 높은 정확도로 분류하는 데 초점을 두고 있으나, 주로 교실 CCTV 또는 데스크톱 환경을 대상으로 하며, 모바일 단말에서의 실시간 처리나 실제 학습 행위 여부 판단은 고려되지 않았다. 또한 자세 정보만으로는 ‘자세는 바르지만 학습하지 않는 상황’을 구분하기 어렵다는 한계가 존재한다.

셋째, 생리신호 기반 연구는 EEG, HRV, GSR 등을 이용해 집중도·정서·인지부하를 측정한다[10]–[12]. 이러한 방식은 헤드셋 등 장비 착용이 필수적이기 때문에 일반 학습 환경에서 적용하기 어렵다.

살펴본 얼굴·시선 기반 접근은 촬영 조건 제약, 자세 기반 접근은 실제 학습 행위 여부 판단 불가, 생리신호 기반 접근은 실용성 부족이라는 한계를 가진다. 무엇보다 기존 연구는 대부분 집중·졸음 등 단일 관찰 지표에 초점을 두어, 학습 행위와 학습 자세를 함께 활용한 학습 행동 인식 체계는 충분히 연구되지 않았다.

본 논문은 기존 고성능 모델들과 달리 스마트폰 단말 환경에서 실시간 처리와 실용성을 목표로 손 모양과 상체 자세 랜드마크를 결합하여 학습자의 고수준 학습 행동 상태를 간결하게 분류한다. 이는 다수의 세분화된 행동을 인식하는 기존 연구와 차별화되는 접근으로, 실제 학습 관리 시스템에 적용 가능한 행동 인식 체계를 제시한다는 점에서 의미를 가진다.

III. 제안하는 학습 행동 인식

3-1 입력 특징 구성 및 모델 학습 환경

본 장에서는 스마트폰 카메라 기반 실시간 학습 행동 인식을 위해 제안하는 인식 모델을 설명한다. 본 논문의 손 모양 인식 모델과 자세 인식 모델은 모두 MediaPipe로부터 추출된 랜드마크 벡터를 입력 특징으로 사용한다. MediaPipe는 각 영상 프레임을 좌표 기반의 정형화된 특징값으로 변환하여 제공한다[13]. 이러한 입력 데이터의 특성은 이미지 전체를 학습하는 CNN 기반 구조나 시간적 의존성을 모델링하는 LSTM·Transformer 계열 모델에 비해 추가적인 공간·시간적 특징 추출 과정이 필요하지 않으므로 본 논문에서는 다층 퍼셉트론(Multi-Layer Perceptron, MLP) 기반 분류 모델을 인식 모델로 사용하였다. 은닉층 깊이(2~4층), 뉴런 수(32~256), Dropout 비율(0.2~0.5) 등 다양한 하이퍼파라미터 조합을 비교하여 모바일 디바이스에서의 실시간 동작을 위해 모델 크기와 추론 속도 등을 고려하여 선택하였다.

모든 학습 실험은 Windows 11 환경에서 Python 3.10, MediaPipe 0.10.5, TensorFlow 2.12.0, Keras 2.12.0을 기반으로 수행되었다. 전 모델은 동일한 전처리 및 정규화(StandardScaler)를 적용하였으며, 전체 데이터셋은 학습 80%, 테스트 20% 비율로 분리하였다. Adam 옵티마이저를 사용하였고, 배치 크기는 32, 에포크는 100으로 설정하였다. 클래스 불균형 문제를 완화하기 위해 클래스별 가중치를 적용하였으며, Dropout을 사용하여 과적합을 방지하였다.

3-2 펜 권손 인식 모델

1) 데이터셋 구축 및 특징 처리

Pen/No Pen 분류에 적합한 공개 데이터셋이 부재함에 따라, 실제 학습 환경을 반영한 자체 촬영 영상을 통해 총 2,145 프레임(Pen 1,035, No Pen 1,110)을 구축하였다. 데이터 수집 시 손 모양의 다양성을 확보하기 위해 촬영 각도, 카메라-손 거리, 조명 조건, 손 크기 등 주요 환경 요소를 변화시키며 다양한 필기 동작을 포함하도록 구성하였다. 본 데이터셋은 제한된 사용자 집단에서 수집되었으나, 손 위치·크기·촬영 조건 변화에 따른 개인차 영향을 완화하기 위한 입력 표현 및 전처리 설계에 중점을 두었다.

실제 환경에서는 손 위치·각도·거리 변화가 빈번하게 발생하므로 모델의 일반화 성능 향상을 위해 가우시안 노이즈, 스케일링, 평면 회전(XY rotation), 평행 이동 등 네 가지 증강 기법을 적용하여 총 21,105프레임 (Pen: 12,075 / No Pen: 9,030)을 확보하였다. 특히 실제 학습 환경에서 발생할 수 있는 스마트폰 거치 각도와 촬영 거리의 가변성을 고려하여, 좌우 $\pm 15^\circ$ 범위의 회전 변환과 0.8~1.2배 스케일 변환을 적용함으로써 다양한 촬영 조건을 모사하였다. 이를 통해 다양한

사용자 환경과 촬영 조건을 모사하며, 펜 캡 여부에 대한 모델의 강건성을 높였다. 이러한 데이터 증강 전략은 특정 사용자나 고정된 촬영 환경에 대한 과적합을 완화하고, 제한된 데이터셋 조건에서도 새로운 사용자 및 환경 변화에 대해 일반화 성능을 확보하기 위한 기술적 보완 방법으로 적용되었다.

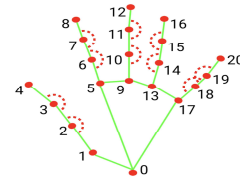


그림 1. 손 랜드마크의 상대좌표 변환 및 사용된 관절각 정의
Fig. 1. Relative coordinate transformation of hand landmarks and joint angles used

MediaPipe Hands를 이용해 각 프레임에서 21개의 손 랜드마크를 추출하였다. 그러나 절대좌표는 손 위치나 시점 변화에 민감하게 반응하여 일반화 성능을 저하시킬 수 있으므로, 손목(랜드마크 0)을 기준으로 모든 좌표를 상대좌표로 변환하였다. 이를 통해 손 위치·크기·촬영 거리 변화에 강인한 사용자 독립적(feature-invariant) 특징 표현을 확보하였다. 이런 상대좌표 기반 표현은 사용자별 손 크기 차이, 카메라 거치 위치 변화, 촬영 시점 차이로 인한 좌표 편차를 효과적으로 정규화하여, 제한된 사용자군에서 학습된 모델이 새로운 사용자 환경에서도 일관된 특징 분포를 유지하도록 한다.

또한 필기 동작을 구조적으로 구분하기 위해 손가락의 굽힘(flexion) 및 마디 사이 각도 변화를 정량적으로 반영할 수 있도록 그림 1과 같이 10개의 관절각(joint angles)을 계산하여 특징 벡터에 포함하였다. 관절각은 손 크기나 절대 위치 변화와 무관하게 필기 동작의 형태적 특성을 반영할 수 있어 개인차 완화에 효과적인 구조적 특징으로 활용된다. 특히 관절각 기반 특징은 동일한 필기 행위가 사용자별로 서로 다른 위치·각도에서 수행되더라도 유사한 패턴으로 표현될 수 있어, 사용자 간 편차를 줄이는 데 핵심적인 역할을 한다.

최종적으로 한 프레임은 73차원 특징 벡터(21개 랜드마크 \times 3축 + 10개 관절각)로 구성되며, 좌표 정규화와 구조적 각도 특징을 결합함으로써 제한된 사용자 데이터 환경에서도 특정 개인이나 촬영 조건에 과적합되지 않고, 필기 동작의 공통 구조를 학습할 수 있도록 설계되었다.

2) 학습 모델

펜 권손 여부를 이진 분류하기 위한 MLP 모델을 설계하였다. 모델 구조는 표 1에 제시한 네 개의 은닉층을 포함하며, 출력층은 이진 분류를 위해 binary_crossentropy 손실 함수와 sigmoid 활성화 함수를 사용하였다.

3) 실험 결과

본 논문에서는 제안한 모델과 어떤 입력 특징 구성이 모바

표 1. 펜 쥐 손 분류 모델 구조

Table 1. Pen-in-hand classification model structure

Layer	Units	Activation function	Layer operations
Input	73 (Relative coordinates + angle features)	-	-
Hidden 1	256	ReLU	BatchNorm + Dropout(0.25)
Hidden 2	128	ReLU	BatchNorm + Dropout(0.25)
Hidden 3	64	ReLU	BatchNorm + Dropout(0.20)
Hidden 4	32	ReLU	Dropout(0.20)
Output	1	Sigmoid	-

일 환경에서의 실시간 손 행위 인식에 가장 효과적으로 기여하는지를 검증하기 위해 세 가지 실험 조건을 비교하였다:

- ① 손모델 1(baseline): 절대좌표 기반 4층 MLP 모델,
- ② 손모델 2: 상대좌표와 관절각을 추가한 4층 MLP 모델,
- ③ 손모델 3: 상대좌표와 관절각을 추가한 5층 MLP 모델.

그림 2는 세 모델의 정밀도-재현율(Precision-Recall, PR) 곡선과 AUC(Area Under Curve)을 보여준다. 손모델 1은 절대좌표 기반 특성으로 인해 위치·시점 변화에 민감하여 PR AUC 0.8216에 머물렀다. 반면 손모델 2는 상대좌표와 관절각을 도입함으로써 구조적 정보를 반영하여 성능이 향상되었으며(PR AUC 0.8912), 손모델 3은 네트워크 깊이를 확장함으로써 전 재현율 구간에서 안정적인 정밀도를 유지하며 가장 높은 PR AUC(0.9699)를 기록하였다. 이는 상대좌표 기반 특징과 모델 구조 확장의 결합이 Pen/No Pen 판별 성능을 효과적으로 향상시킴을 보여준다. 그림 3은 제안한 손모델 3이 실제 환경에서 펜을 쥐 손과 그렇지 않은 손을 안정적으로 구분한 예시를 나타낸다.

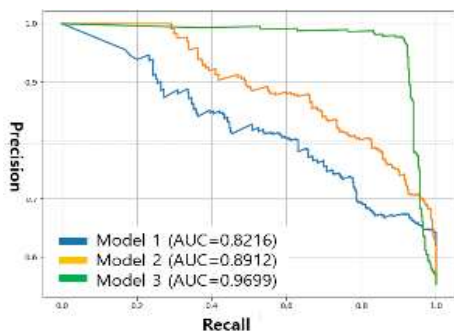


그림 2. 펜 쥐 손 학습 모델의 정밀도-재현율 곡선
Fig. 2. Precision-recall curve of the pen-in hand classification model

표 2는 최종 선택된 손모델 3의 혼동 행렬을 나타내며, 전체 정확도 93.7%를 기록하였다. False Negative가 29로 낮게 나타나 펜을 쥐 손을 미검출하는 사례가 감소하여 실제 학습 행위 인식의 신뢰성이 향상되었음을 확인하였다.

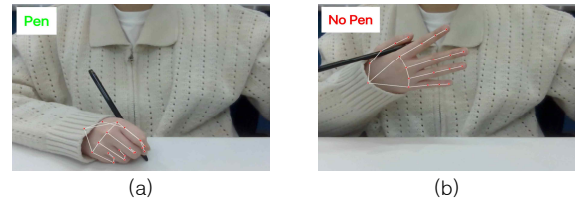


그림 3. 손모델 3으로 인식된 펜 쥐 손 인식 예시
Fig. 3. Example results of pen-in hand classification using hand model 3: (a) Pen, (b) No pen

표 2. 손 모델3의 펜 쥐 손 인식 혼동 행렬

Table 2. Confusion matrix of the pen-in hand classification using hand model 3

		Predicted	
		No Pen	Pen
Actual	No Pen	291	14
	Pen	29	348

표 3. 손 모델3의 LOSO 교차 검증 결과

Table 3. LOSO cross-validation results using hand model 3

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Accuracy (%)	82.3	80.7	81.8	82.8	80.4

또한 제안한 손 인식 모델의 실제 적용 가능성을 검증하기 위해 학습에 참여하지 않은 사용자를 대상으로 Leave-One-Subject-Out(LOSO) 교차 검증을 수행하였다. 총 5명의 사용자에 대해 한 명을 테스트 세트로 분리하고 나머지 사용자 데이터로 학습하였으며, 각 fold에서는 No Pen 1,806프레임과 Pen 2,415프레임을 사용하였다. 그 결과 손모델 3은 표 3과 같이 fold별 80.4%~82.8%의 정확도를 기록하였고, 평균 정확도는 81.6%로 나타났다. 이는 무작위 분할 평가 대비 개인 간 변동성으로 인해 약 12.1%p의 성능 감소가 발생한 것으로, 기존 HAR 연구에서 보고된 사용자 독립 평가에서 일반적으로 관찰되는 수준이며[2], 제한된 사용자 수 환경에서도 제안 모델이 특정 사용자에게 과도하게 의존하지 않고 안정적인 판별 성능을 유지함을 시사한다.

3-3 공부 자세 인식




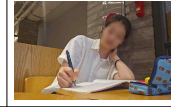
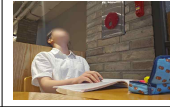

1) 공부 자세에 대한 정의

학습자의 공부 자세는 신체적 안정성과 집중도 유지에 직접적인 영향을 미치는 요소로, 교육학·운동과학·HCI 분야에서 지속적으로 다루어져 온 핵심 연구 주제이다. 선행 연구에서는 자세 인식 기술이 다양한 인간 행동 분류 및 상태 추정 의 핵심 지표로 활용되어 왔으며[2], 이는 학습 상황에서도 상체 자세가 사용자의 신체적·심리적 상태를 유의미하게 반영할 수 있음을 시사한다.

본 논문에서는 손 모양을 이용한 학습 행위 인식과는 별도로, 학습자의 자세를 판단하기 위한 6가지 공부 자세 유형을 정의한다(표 4). 이 6가지 자세는 상체의 정렬과 기울기 패턴

표 4. 6가지 공부 자세 정의

Table 4. 6 study postures according to study states

Study states	(a) Proper study state	(b) Low concentration state			(c) Sleep state	
Defined postures	(1) A neutral spine posture with balanced shoulder heights	(2) A posture where the sit bones are positioned near the front of the chair and the upper body is tilted backwards	(3) A posture where the neck is bent forward and the upper body is supported on a desk or arm	(4) A posture where the neck and shoulders are bent to one side and the body is supported with one arm or hand	(5) A posture in which the upper body is tilted backward and the neck joint angle exceeds 180°, causing the head to fall behind the chair.	(6) A posture in which the head is bent forward and the face is not visible, making study activity difficult to determine.
Label	good_pose	nfocus_lean_back	nfocus_lean_forward	nfocus_lean_side	sleep_head_back	sleep_head_down
Raw data #	1,578	1,160	1,434	724	710	538
Augmentated data #	3,200	3,200	3,200	3,200	3,200	3,200
Example image						

에 따라 (a) 바른 공부 상태, (b) 집중도 저하 상태, (c) 수면 상태의 세 가지 범주로 구분된다. 바른 공부 상태(a)는 척추 중립과 어깨 수평 정렬이 유지된 안정적인 상태를 의미하며, 집중도 저하 상태(b)는 상체가 뒤로 기울어지거나 앞으로 숙여지거나 좌·우 비대칭으로 치우치는 등 신체적 불완전성이 증가한 상태를 포함한다. 수면 상태(c)는 머리가 뒤로 젖혀지거나 상체가 전방으로 굴곡되어 얼굴이 보이지 않는 등의 수면 특성을 보이는 자세로 정의한다.

2) 데이터셋 구축 및 특징 처리

본 논문에서는 학습자의 상체 자세를 기반으로 공부 자세를 분류하기 위한 데이터셋을 직접 구축하였다. 기존 공개 데이터셋은 일반적인 인체 동작 인식에 초점을 두고 있어, 학습 상황에서 나타나는 미세한 상체 기울기, 머리 기울기, 줄음 상태와 같은 세부 차이를 구분하기에는 적합하지 않다. 이에 논문에서 정의한 6가지 상체 자세를 실제 스마트폰 기반 학습 관리 시스템에서 빈번하게 사용될 것으로 가정한 촬영 구도를 중심으로 실제 학습 환경과 유사한 조건을 구성하여 영상을 촬영하였다. 촬영은 책상 위에 고정된 스마트폰 또는 카메라가 학습자를 내려다보는 형태를 반영하기 위해 스마트폰 거치 환경에서 일반적으로 사용되는 시야를 기준으로 약 30~60° 범위의 하향 각도, 일정 조도, FHD 해상도, 30 FPS 조건에서 진행되었으며, 촬영된 영상은 프레임 단위로 추출하였다. 이는 다양한 카메라 배치 중 본 연구 시스템의 실제 사용 시나리오를 대표하는 조건으로 설정된 것이다. 비례당 자세 구간은 수동으로 제거하여 라벨의 정확성을 확보하였다. 전체 원본 데이터는 6,184프레임으로 구성되며, 자세별 분포는 표 4에 제시하였다.

클래스 불균형 완화와 다양한 환경 변화를 모사하기 위해 회전 증강과 좌우 반전 증강을 적용하였다. 상체 자세 데이터에 대한 증강은 앞서 손 데이터 증강에서 설명한 촬영 각도·

환경 변화 대응 전략과 일관되게 설계되었으며, 상체와 얼굴 방향 변화에 초점을 두어 적용되었다. 회전 증강은 10°~15° 범위의 랜덤 회전을 통해 상체·얼굴 방향 변화를 학습하도록 하였으며, 좌우 반전 증강은 인체 대칭을 고려해 x 좌표를 반전하는 방식으로 수행하였다. 이를 통해 특정 촬영 각도나 자세 편향에 대한 의존성을 줄이고, 제한된 촬영 환경에서도 다양한 상체 자세 변화를 안정적으로 학습할 수 있도록 하였다. 증강 후 각 클래스는 균등하게 3,200개로 확대되었고 전체 데이터는 19,200개로 증가하여 모델의 일반화 성능 향상에 기여하였다.

MediaPipe Pose로 상체 주요 관절 12개(랜드마크 11~22), MediaPipe FaceMesh로 얼굴 구조적 특징을 반영하는 19개 포인트를 추출하였다. 모든 프레임은 초당 3프레임으로 표본화하였고, 각 랜드마크는 (x, y, z) 좌표로 구성하였다. 상체 랜드마크는 전신이 아닌 어깨·팔·몸통 중심 관절로 한정함으로써, 다리 위치나 하반신 움직임과 같이 학습 행동 판단에 직접적으로 기여하지 않는 변수를 제거하고, 착석 학습 환경에서 반복적으로 관찰되는 상체 정렬·기울기 패턴에 집중하도록 설계하였다.

전처리 과정에서는 얼굴이 가려져 FaceMesh 검출이 실패한 경우 얼굴 좌표를 0으로 대체해 얼굴을 숙이고 수면하는 자세(sleep_head_down)의 특성을 반영하였고, Pose와 FaceMesh가 동시에 실패한 프레임은 제거하였다. 이와 같은 처리 방식은 조명 변화, 부분 가림, 고개 숙임 등 실제 학습 환경에서 빈번하게 발생하는 얼굴 검출 실패 상황을 하나의 의미 있는 자세 신호로 해석함으로써, 특정 사용자나 촬영 조건에 의존하지 않는 강건한 상태 표현을 가능하게 한다.

또한 일부 영상에서 발생하는 상하 반전 문제를 해결하기 위해 수직 플립 보정을 수행하였다. 이는 스마트폰 거치 방향이나 카메라 설치 방식 차이로 인해 발생할 수 있는 좌표 방향성 불일치를 완화하여, 동일한 학습 자세가 서로 다른 촬영

조건에서도 일관되게 표현되도록 하기 위한 조치이다.

최종적으로 모든 샘플은 Pose 36차원과 FaceMesh 57차원을 합한 총 93차원 특징 벡터로 구성되었다. 상체 골격 정보와 얼굴 기울기·하향 여부를 함께 활용함으로써, 체형이나 얼굴 크기 차이보다는 자세 변화의 상대적 패턴에 기반한 분류가 가능하도록 하여 사용자 간 편차에 대한 민감도를 완화하였다. 다만 본 데이터셋은 특정 촬영 각도 범위를 전제로 구축되었으며, 이는 스마트폰을 책상 위에 거치하는 일반적인 학습 환경을 반영한 설정이지만, 정면·측면 등 다양한 카메라 배치 환경에 대한 일반화 검증은 향후 연구 과제로 남는다.

3) 학습 모델

공부 자세를 분류하기 위한 MLP 모델을 설계하였다. 모델 구조는 표 5에 제시한 네 개의 은닉층을 포함하며, 6가지 출력 값을 위해 sparse_categorical_crossentropy 손실함수와 softmax 활성화 함수를 적용하였다.

표 5. 공부 자세 분류 모델 구조

Table 5. Study posture classification model structure

Layer	Units	Activation function	Layer operations
Input	93 (face and upper body)	-	-
Hidden 2	128	ReLU	BatchNorm + Dropout(0.3)
Hidden 3	64	ReLU	BatchNorm + Dropout(0.3)
Hidden 4	32	ReLU	Dropout(0.3)
Output	1	Softmax	-

4) 실험 결과

본 논문에서는 네트워크 깊이는 고정하고 다음 세 가지 입력 특징 구성을 비교하여 상체 자세 분류에 가장 효과적인 조합을 분석하였다:

- ① 자세모델 1(baseline): Pose 좌표 (0-22)만 사용
- ② 자세모델 2: Pose 좌표 + Head-Shoulder 각도
- ③ 자세모델 3: Pose (11-22) + FaceMesh 19개 특징점

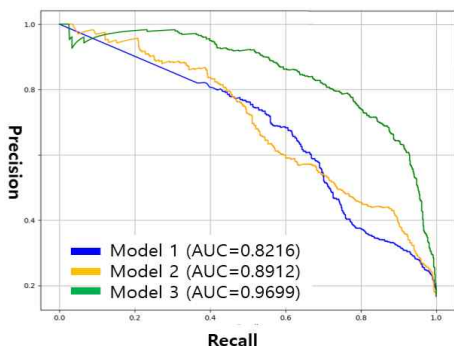


그림 4. 공부 자세 학습 모델의 정밀도-재현율 곡선
Fig. 4. Precision-recall curve of the study posture classification model

그림 4는 세 자세모델의 PR 곡선과 AUC를 보여준다. 상체 랜드마크(0-22)를 사용하는 자세모델 1은 공부 자세 인식을 위한 기준 모델로, PR 곡선에서도 재현율이 증가할 때 정밀도가 급격히 하락하는 불안정한 패턴이 확인되었다. 이는 Pose 좌표만으로는 촬영 거리, 기울기 변화, 사용자 체형 차이에 따른 외부 요인을 효과적으로 보정하기 어렵기 때문이다. 자세모델 2는 Head-Shoulder 각도를 추가하여 기울기 정보를 보강하였으나, 특징 공간 복잡성 증가로 정밀도-재현율 곡선에서도 전반적인 추가 효과가 제한적으로 나타났다.

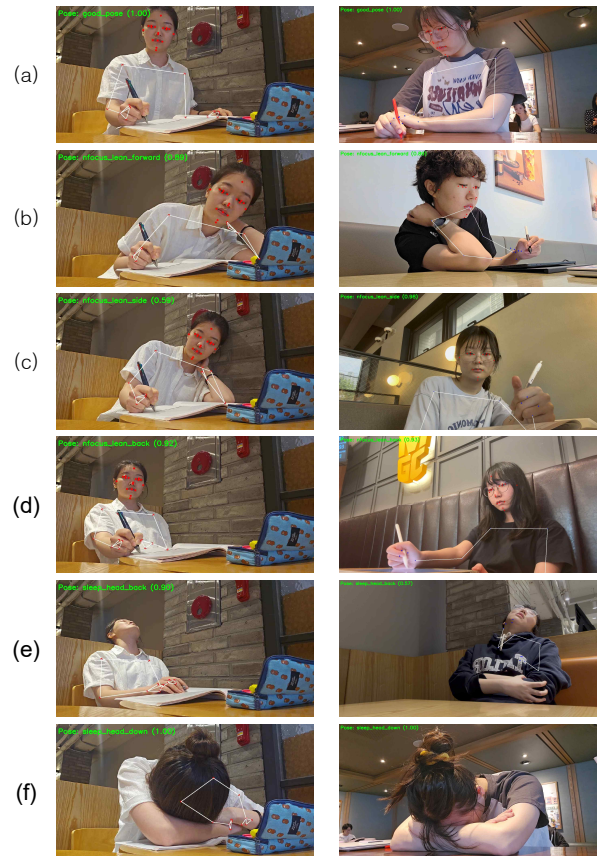


그림 5. 자세모델 3으로 인식된 공부 자세 예시
Fig. 5. Example results of study posture classification using posture model 3: (a) good_pose, (b) nfocus_lean_forward, (c) nfocus_lean_side, (d) nfocus_lean_back, (e) sleep_head_back, and (f) sleep_head_down

반면, 자세모델 3은 MediaPipe Pose에서 상체 중심 랜드마크(11-22)와 훨씬 더 촘촘한 얼굴 구조 정보를 제공하는 FaceMesh의 19개 랜드마크를 결합하여 얼굴 전후·좌우 기울기와 시선 방향 차이 등 자세 판별 핵심 정보를 보강하였다. 이 결과 PR AUC는 0.9699로 가장 높은 성능을 나타냈으며, 재현율도 80% 이상의 성능을 보였다. 이는 전후·좌우 기울기 및 수면 자세 구분이 어려운 상황에서도 얼굴 방향 정보가 보완적 역할을 수행한 것으로 해석된다. 그림 5는 최종 선택된

자세모델 3의 실제 인식 예시를 제공한다.

자세모델 3의 혼동행렬(표 6)에 따르면 전체 정확도는 약 84%로 나타났으며, 대부분의 자세가 안정적으로 분류되었다. 일부 lean_side 자세는 45° 하향 촬영 환경에서 좌우-전후 기울기가 유사한 좌표 패턴으로 표현되는 MediaPipe Pose의 구조적 특성과 lean_side 자세 자체의 높은 자세 변동성으로 인해 혼동이 발생하였다. 그럼에도 lean_side 자세는 최종 학습 행동 인식 단계에서 ‘집중력 저하’로 통합되므로, 자세모델 3의 학습 행동 인식 성능에 미치는 영향은 제한적이다.

표 6. 자세모델3의 공부 자세 인식 혼동 행렬

Table 6. Confusion matrix of the study posture classification using posture model 3

		Predicted					
		good pose	lean back	lean forward	lean side	head back	head down
Actual	good pose	130	0	0	5	0	0
	lean back	0	114	0	20	0	2
	lean forward	0	0	101	2	0	0
	lean side	13	9	29	67	0	0
	head back	0	0	0	5	80	6
	head down	0	0	0	3	11	58

3-4 통합 학습 행동 인식

1) 학습 행동(Behavior)의 정의

본 논문에서는 학습자의 실제 학습 수행 여부와 학습 태도 변화를 구분하기 위해 네 가지 학습 행동 상태—‘집중’, ‘집중력 저하’, ‘수면’, ‘공부 중단’—를 정의한다. 이 중 ‘집중’과 ‘집중력 저하’는 학습이 수행 중인 상태에서 태도의 질적 차이를 나타내며, ‘수면’은 학습 수행과 명확히 구별되는 비활동 상태로 정의된다[14]. ‘공부 중단’은 수면 상태가 아니면서 학습 행위의 직접적 지표인 필기 동작이 관찰되지 않는 경우로 설정하였다. 이는 본 시스템이 펜 보유 여부를 학습 수행 판단의 핵심 기준으로 사용한다는 설계 특성을 반영한다. 다만 필기를 수반하지 않는 학습 활동을 구분하지 못하는 한계가 있으며, 이는 향후 확장이 필요한 부분이다.

이 네 가지 상태는 상호 배타적으로 정의되며, 실제 학습 환경에서 발생하는 주요 행동 유형을 단순하면서도 명확하게 구분하는 것을 목표로 한다. 표 7은 그림 6의 통합 절차를 기반으로 손 모양 인식 결과와 자세 분류 결과를 조합해 최종 학습 행동을 결정하는 규칙 기반 의사결정 알고리즘을 의사코드(pseudocode) 형태로 정리한 것이다.

2) 통합 학습 행동 인식 알고리즘

본 연구의 통합 모델은 손 모양 인식 결과(Pen/No Pen)와 상체 자세 분류 결과(6개 자세)를 규칙 기반으로 결합하여 최종 학습 행동을 결정한다. 단일 모델만 사용할 경우 발생하는 모호성을 해소하는 것이 통합 알고리즘의 핵심 목적이다. 예

표 7. 학습 행동 상태 통합 알고리즘(Pseudocode)

Table 7. Integrated study behavior state classification algorithm(pseudocode)

```

Input: hand_state ∈ {PEN, NO_PEN}
       posture_state ∈ {GOOD, LEAN_BACK, LEAN_FORWARD,
                       LEAN_SIDE, SLEEP_BACK, SLEEP_DOWN}
if posture_state ∈ {SLEEP_BACK, SLEEP_DOWN} then
    behavior ← SLEEP
else if hand_state = NO_PEN then
    behavior ← STOP_STUDYING
else
    if posture_state = GOOD then
        behavior ← FOCUS
    else
        behavior ← REDUCED_FOCUS
    end if
end if
return behavior
    
```

를 들어, 펜을 쥐고 있으나 자세가 무너진 경우는 단순 학습 수행 상태가 아니라 집중 저하로 해석되어야 하며, 반대로 펜을 쥐지 않은 경우에는 학습 중단과 수면 상태를 자세 정보를 통해 구분할 필요가 있다.

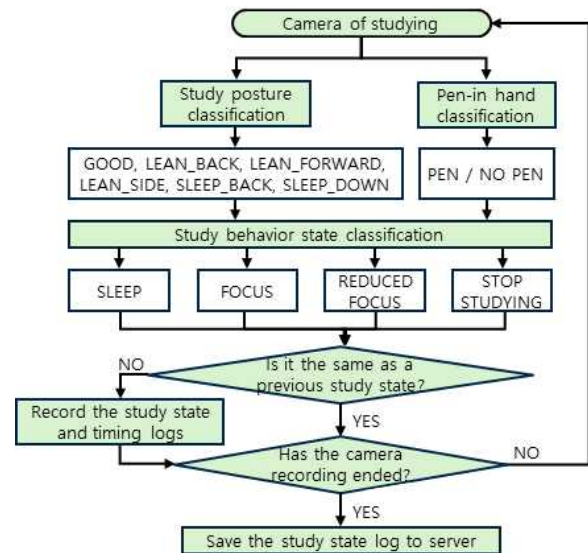


그림 6. 학습 행동 인식을 위한 통합 절차도

Fig. 6. Flowchart of the integrated study behavior classification process

그림 6의 절차에 따라, 통합 알고리즘은 우선 수면 관련 자세를 탐지하여 ‘수면’으로 분류한다. 수면이 아닌 경우에는 펜을 쥐는 상태에서만 자세 분류 결과에 따라 ‘집중’ 또는 ‘집중력 저하’를 결정하며, 펜을 쥐지 않은 상태에서는 ‘공부 중단’으로 판단한다. 이러한 규칙 기반 통합은 각 단일 모델의 역할을 명확히 분리하면서도, 실제 학습 장면에서 발생하는 핵심 행동 패턴을 효율적으로 통합할 수 있도록 설계되었다.

시스템은 실시간 영상 입력을 처리하여 상태 변화가 발생한 시점에만 로그를 생성하고, 학습 종료 후 모든 기록을 서버에 저장해 이후 분석에 활용한다.

3) 실험 결과

표 8은 손 모양 인식 모델과 자세 인식 모델의 출력을 통합하여 네 가지 학습 행동 상태로 분류한 결과를 정리한 것이다. 통합 모델은 전체 정확도 86.2%, 평균 재현율 86%를 기록하여, 단일 모델만으로는 구분이 어려운 학습 행동 상태를 효과적으로 분리할 수 있음을 확인하였다.

그림 7의 PR 곡선은 네 가지 학습 행동 모두에서 안정적인 분류 특성을 보여 통합 모델의 일관된 분류 성능을 확인시켜 준다. 또한 표 9의 혼동 행렬 분석 결과, ‘집중’과 ‘집중력 저하’, ‘수면’, ‘공부 중단’ 상태가 안정적으로 구분되었으며, 일부 경계 상황을 제외하고 특정 상태로의 편향된 오분류는 제한적으로 나타났다. 이는 손 모양 기반 학습 수행 판단과 자세 기반 태도 분류를 결합함으로써, 단일 신호 기반 접근이나 시간 측정 중심 학습 관리 방식보다 실제 학습 행동을 보다 일관되고 안정적으로 분류할 수 있음을 실험적으로 입증한다.

표 8. 통합 학습 행동 상태 인식 정확도

Table 8. Classification accuracy of the integrated study behavior state classification model

Study behavior	precision	recall	f1-score
Focus	0.63	0.83	0.72
Reduced Focus	0.85	0.89	0.87
Sleep	0.99	0.94	0.96
Stop Studying	0.89	0.78	0.83

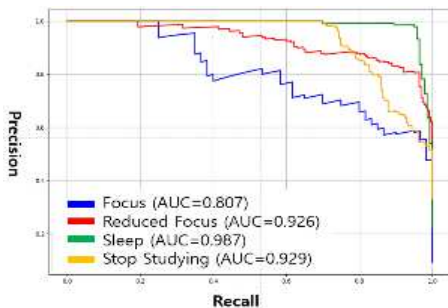


그림 7. 학습 행동 상태별 정밀도-재현율 곡선

Fig. 7. Precision-recall curves of the 4 study behavior states

표 9. 최종 통합 학습 행동 인식의 혼동 행렬

Table 9. Confusion matrix of the integrated study behavior classification model

		Predicted			
		Focus	Reduced Focus	Sleep	Stop Studying
Actual	Focus	50	3	0	7
	Reduced Focus	11	203	1	12
	Sleep	0	6	135	2
	Stop Studying	18	28	1	168

표 10은 영상 인식 기반 학습자 행동 및 자세 인식 연구들과 본 논문의 접근 방식을 비교한 것이다. 기존 연구들은

CNN 또는 어텐션 기반 고성능 모델을 활용하여 다수의 세분화된 행동을 인식하며, 교실 CCTV나 GPU 기반 환경을 주요 적용 대상으로 한다. 반면, 본 논문은 스마트폰 단말에서 교사 양 연산 자원 없이 실시간 구동이 가능하도록 설계된 경량 구조를 기반으로, 손 모양과 상체 자세 랜드마크를 결합해 학습자의 고수준 학습 행동 상태를 간결하게 분류한다는 점에서 실용적 차별성을 가진다. 따라서 표 10의 비교는 절대적인 정확도 성능의 우열을 판단하기보다는, 적용 환경과 문제 정의의 차이에 따른 연구 방향과 시스템 설계 관점의 차이를 명확히 하기 위한 목적으로 제시된다.

IV. Studify: 제안하는 학습 상태 자동 기록 및 피드백 시스템

본 장에서는 제안한 학습 행동 인식 모델을 기반으로 구현한 모바일 학습 관리 시스템 Studify의 구조와 기능을 설명한다. Studify는 스마트폰 카메라로부터 펜을 쥔 손 모양과 상체 자세를 실시간으로 분류하여 네 가지 학습 행동 상태를 자동 기록하고, 서버와 연동해 학습 상태 분석 결과와 맞춤형 피드백을 제공하도록 설계되었다.

4-1 시스템 개요 및 전체 아키텍처

Studify는 Android 클라이언트, Spring Boot 서버, MySQL 데이터베이스로 구성된 3계층 구조를 기반으로 한다. 그림 8은 시스템 전체 구조를 나타낸다. 클라이언트는 카메라 프레임을 기반으로 MediaPipe Hands, FaceMesh, Pose를 적용하여 손·상체 랜드마크를 추출하고, 이를 TensorFlow Lite 모델을 통해 실시간으로 학습 행동 상태를 분류한다. 분석된 상태 정보는 학습 종료 시 서버로 전송된다.

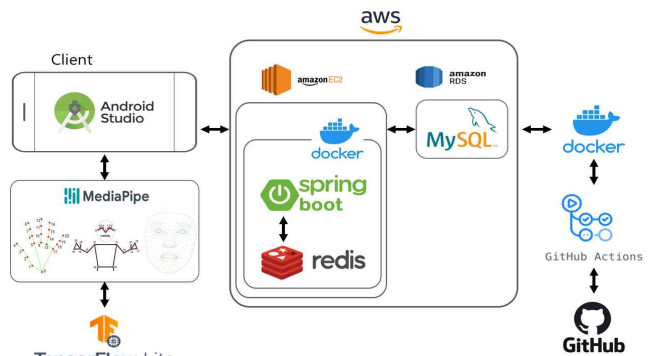


그림 8. Studify 시스템 구조도
Fig. 8. System architecture of Studify

서버는 사용자 정보·학습 기록을 저장하고 통계 요청을 처리하며, GPT 기반 피드백 생성을 위한 API를 수행한다. 전체 시스템은 Docker 기반으로 배포되고 GitHub Actions로 자

표 10. 영상 인식 기반 학습자 행동·자세 인식 연구 비교

Table 10. Comparison with related vision-based student behavior/posture recognition studies

Study (Year)	Model / Architecture	Target Behavior Level	Key Characteristics	Performance
[7] (2021)	Improved HRNet	Fine-grained posture	CNN, full-body pose, classroom CCTV	mAP 73.76%
[8] (2025)	EPLC-Pose	Upper-body posture	Lightweight CNN, edge-oriented	Higher than baselines
[9] (2024)	AdaptSepCX(Attention-based)	Student actions (online)	Attention model, online learning	Improved over CNN
Proposed Method	MediaPipe + 2 MLPs + rule-based fusion	High-level study behavior	Mobile real-time, landmark-based, no extra device	Accuracy 86.2%

동 빌드·배포되어 운영 환경의 일관성을 유지한다. 데이터베이스는 AWS RDS(MySQL)를 사용하며, Redis는 인증 및 세션 캐싱 처리에 활용된다.

4-2 학습 행동 상태 자동 기록 기능

id	end_time	start_time	state	study_record_id
2782	2025-10-31 15:50:11.000000	2025-10-31 15:49:41.000000	STUDYING	304
2783	2025-10-31 15:50:32.000000	2025-10-31 15:50:20.000000	STUDYING	304
2784	2025-10-31 15:49:57.000000	2025-10-31 15:49:51.000000	GOOD_POSE	304
2785	2025-10-31 15:50:11.000000	2025-10-31 15:50:00.000000	GOOD_POSE	304
2786	2025-10-31 15:50:25.000000	2025-10-31 15:50:20.000000	GOOD_POSE	304
2787	2025-10-31 15:50:32.000000	2025-10-31 15:50:29.000000	GOOD_POSE	304
2788	2025-10-31 15:49:51.000000	2025-10-31 15:49:41.000000	NFOCUS_POSE	304
2789	2025-10-31 15:50:00.000000	2025-10-31 15:49:57.000000	NFOCUS_POSE	304
2790	2025-10-31 15:50:29.000000	2025-10-31 15:50:25.000000	NFOCUS_POSE	304
2791	2025-10-31 15:50:47.000000	2025-10-31 15:50:41.000000	SLEEP_POSE	304
2792	2025-10-31 15:50:41.000000	2025-10-31 15:50:34.000000	AWAY	304
2793	2025-10-31 15:49:41.000000	2025-10-31 15:49:38.000000	ETC	304
2794	2025-10-31 15:50:20.000000	2025-10-31 15:50:11.000000	ETC	304
2795	2025-10-31 15:50:34.000000	2025-10-31 15:50:32.000000	ETC	304

그림 9. 학습 행동 상태 인식 로그 데이터 예시

Fig. 9. Example of logged study behavior states during a study session

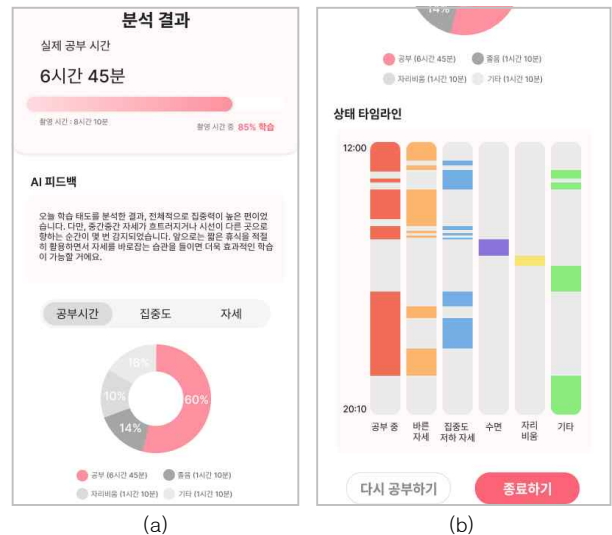
Studify 시스템은 기존 타이머 기반 기록 방식과 달리, 학습 모델을 통해 사용자의 학습 상태를 자동 분류한다. 사용자가 ‘공부 시작’을 누르면 시스템은 일정 간격으로 영상을 분석해 ‘집중’, ‘집중도 저하’, ‘수면’, ‘공부 중지’의 네 가지 상태를 인식한다. 또한 매 프레임이 아닌 행동 상태 변화가 발생한 시점에만 새로운 로그를 기록하여 효율적으로 학습 타임라인을 구성한다. 그림 9는 실제 기록된 로그의 예시로, 각 로그는 상태 시작·종료 시각과 해당 행동 상태로 이루어진다. 이를 기반으로 총 학습 시간, 집중 유지 시간, 수면 시간, 자리 비움 시간이 자동 산출된다. 분석 결과는 타임라인과 상태 비율 그래프로 제공되어 사용자의 학습 패턴 분석을 지원한다.

4-3 분석 결과 시각화 및 개인 맞춤형 AI 피드백 기능

Studify는 단순한 학습 시간 기록 도구를 넘어, 사용자 행동 데이터를 기반으로 한 정량적 분석과 개인화된 인지·행동 지원 기능을 제공하도록 설계되었다. 그림 10(a)는 학습 세션 동안의 집중 시간, 자세별 비율, 수면 및 자리 비움 시간 등을 집계한 분석 요약 화면으로, 사용자가 자신의 학습 패턴을 종합적으로 파악할 수 있도록 구성되어 있다. 그림 10(b)는 학습 중 발생한 상태 변화를 시간 흐름에 따라 시각화한 타임라인으로, 특정 시점에서의 집중 유지 여부와 집중 저하·수면

발생 구간을 직관적으로 확인할 수 있다.

또한 Studify는 이러한 행동 로그를 기반으로 GPT API를 활용한 개인 맞춤형 피드백 기능을 제공한다. 서버는 사용자의 집중·비집중·수면·공부 중단 지속 시간 및 비율, 자세 변화 패턴 등을 요약하여 GPT 모델의 입력으로 전달하고, 모델은 이를 바탕으로 학습 태도 진단과 개선 전략을 포함한 자연어 피드백을 생성한다. 예를 들어 집중 저하 구간이 길거나 졸음 상태가 반복적으로 감지된 경우, 시스템은 “짧은 집중 구간과 규칙적 휴식을 병행하는 학습 전략을 적용할 것”, “학습 환경을 정돈하여 자세 흐트러짐을 유발하는 요인을 줄일 것”과 같은 구체적 조언을 제공한다.



These figures are screen captures of the developed application, and this application is made for Korean users.

그림 10. 분석 결과 앱 화면 예시: (a) 분석 결과 요약, (b) 학습 타임라인 시각화

Fig. 10. App screenshots of analysis results: (a) Summary of analysis results, (b) visualization of the study timeline

이와 같은 시각화 및 피드백 기능은 Studify가 단순 시간 측정 도구가 아니라, 학습자의 행동 데이터를 기반으로 한 정량적 분석과 개인화된 조언을 통합하는 학습 지원 시스템으로 기능하도록 하는 핵심 요소라 할 수 있다. 이를 통해 사용자는 자신의 학습 습관을 객관적으로 이해하고, 행동 기반 개선 전략을 지속적으로 적용할 수 있는 환경을 제공받는다.

4-4 실험 결과: 사용자 만족도 평가

본 논문에서 제안한 스마트폰 기반 학습 행동 자동 기록 시스템의 실사용 가능성과 사용자 수용도를 검증하기 위해 총 26명의 대학생 및 일반 사용자를 대상으로 사용자 만족도 조사를 실시하였다. 참여자들은 약 5-10분간 애플리케이션을 사용한 후, 학습 보조 유용성, 핵심 기능에 대한 만족도, 통계 및 피드백 기능 평가, 지속 사용 의향 등 총 5개 문항에 대해 5점 리커트 척도로 응답하였다.

표 11은 설문 항목별 평균 점수를 나타낸다. 응답자들은 전반적으로 제안 시스템의 핵심 기능에 대해 높은 만족도를 보였다. 특히 펜 퀴 손 인식 기반 실제 공부 시간 측정의 정확성, 자세 인식 기반 집중도 추정 기능의 신기함과 창의성, 자세 교정을 통한 학습 습관 개선 효과, 자리 비움 자동 감지 등 자동화 기능의 편의성, 직관적인 UI·디자인 요소가 긍정적으로 평가되었다. 이는 제안 시스템이 기존 타이머 기반 방식과 비교해 학습자의 실제 행동을 정밀하게 반영하고, 학습 과정에 자연스럽게 통합될 수 있음을 시사한다. 한편, 일부 응답자는 서서 조는 자세 등 추가적인 자세 인식 확장, 다크모드 제공, 업적·포인트 기반 동기 부여 기능, 측정 준비 과정의 간소화 등 개선점을 제안하였다. 이는 전반적인 사용성은 높으나 다양한 사용 환경 대응력과 사용자 경험(UX) 향상을 위한 기능적 확장이 향후 시스템 개선 방향으로 요구됨을 의미한다.

표 11. 사용자 만족도 조사(5점 척도)

Table 11. User satisfaction survey(5-point Likert scale)

Survey items	Mean
Usefulness for supporting real study activities	4.96
Satisfaction with pen-in-hand-based time measurement	4.96
Satisfaction with posture-based concentration estimation	4.87
Satisfaction with statistics and feedback features	4.88
Willingness to continue using the application	4.96

V. 결 론

본 논문은 스마트폰 카메라 기반으로 손 모양과 상체 자세를 통합 분석하여 학습자의 실제 학습 행동을 자동 분류하는 모델을 제안하였다. MediaPipe에서 추출된 손·자세·얼굴 랜드마크를 입력으로 하는 두 개의 MLP 모델—펜 퀴 손 인식 모델과 상체 자세 분류 모델—을 설계하고, 모바일 환경에서의 실시간 처리를 고려하여 특정 구성과 네트워크 깊이에 따른 성능 변화를 비교하였다. 상대좌표 및 관절각을 사용한 손 인식 모델은 93.7%의 정확도를, 상체 Pose와 FaceMesh 특징을 결합한 자세 모델은 84%의 정확도를 기록하였다. 또한, 두 모델의 출력을 규칙 기반으로 통합한 학습 행동 분류는 네 가지 행동 상태(집중, 집중력 저하, 수면, 공부 중단)를 86.2%의 정확도로 분류하여 동일 데이터 조건에서 단일 모

델 대비 보다 일관된 학습 행동 구분이 가능함을 확인하였다.

제안된 인식 모델은 모바일 애플리케이션 Studify에 적용되어 실시간 행동 기록, 타임라인 시각화 및 AI 기반 피드백 기능을 제공한다. 사용자 평가 결과는 자동 기록 기능의 편의성, 학습 상태 분석의 유용성, 피드백의 실질적 도움 측면에서 높은 만족도를 확인하였다.

다만 본 연구는 제한된 사용자군과 스마트폰을 책상 위에 거치한 30-60° 하향 각도 환경을 전제로 설계·평가되었기 때문에, 성별·연령·손 크기·체형 차이 등 개인별 편차와 다양한 스마트폰 거치 환경에 대한 일반화 성능은 충분히 검증되지 않았다. 또한 학습 행동 판단이 필기 행위 여부에 기반함에 따라, 독서·타이핑·온라인 강의 수강과 같은 비필기 학습 활동을 명확히 구분하지 못하는 구조적 한계를 지닌다.

향후 연구에서는 다양한 성별·연령·체형을 포함한 사용자 집단과 카메라 각도·거리·조명 조건이 상이한 환경에서의 데이터 확장을 통해 제안 모델의 일반화 성능을 정량적으로 분석할 예정이다. 특히 LOSO 검증과 같은 사용자 독립 평가를 상체 자세 모델 및 통합 행동 인식 단계로 확장하여, 새로운 사용자에 대한 성능 유지 여부를 체계적으로 검증하고자 한다. 아울러 본 연구에서 권장한 스마트폰 하향 거치 각도뿐 아니라, 태블릿 거치 환경이나 노트북 상단 카메라와 같은 다양한 학습 기기 및 거치 조건으로 적용 범위를 확장함으로써, 실제 학습 환경 전반에서 활용 가능한 범용적 학습 행동 인식 체계로 발전시키는 것을 목표로 한다.

참고문헌

- [1] OECD, Education at a Glance 2024, OECD Indicators, OECD Publishing, Paris, 2024. <https://doi.org/10.1787/c00cad36-en>
- [2] M. Nadeem, E. Elbasi, A. I. Zreikat, and M. Sharsheer, "Sitting Posture Recognition Systems: Comprehensive Literature Review and Analysis," *Applied Sciences*, Vol. 14, No. 18, 8857, 2024. <https://doi.org/10.3390/app14188557>
- [3] D. Xie, X. Zhang, X. Gao, H. Zhao, and D. Du, "MAF-Net: A Multimodal Data Fusion Approach for Human Action Recognition," *PLoS One*, Vol. 20, No. 4, e0319656 2025. <https://doi.org/10.1371/journal.pone.0319656>
- [4] T. Robal, Y. Zhao, C. Lofi, and C. Hauff, "Webcam-Based Attention Tracking in Online Learning: A Feasibility Study," in *Proceedings of the 23rd International Conference on Intelligent User Interfaces*, Tokyo: Japan, pp. 189-197, 2018. <https://doi.org/10.1145/3172944.3172987>
- [5] R. Daza, L. F. Gomez, J. Fierrez, A. Morales, R. Tolosana, and J. Ortega-Garcia, "DeepFace-Attention: Multimodal Face Biometrics for Attention Estimation with Application to e-Learning," *IEEE Access*, Vol. 12, pp. 111343-111359,

2024. <https://doi.org/10.1109/ACCESS.2024.3437291>

- [6] X. Hu and J. Gao, "Facial Expression Recognition Reveals Students' Engagement in Online Class: Correlations with Six Engagement Measurements," *PLoS ONE*, Vol. 20, No. 10, e0334232, 2025. <https://doi.org/10.1371/journal.pone.0334232>
- [7] Y. Zhang, T. Zhu, H. Ning, and Z. Liu, "Classroom Student Posture Recognition Based on an Improved High-Resolution Network," *Journal of Wireless Communications and Networking*, Vol. 2021, 140, 2021. <https://doi.org/10.1186/s13638-021-02015-0>
- [8] Y. Ji, Y. Jin, Z. Wang, and S. Tan, "EPLC-Pose: A Lightweight Student Posture Recognition Network Under Panoramic Classroom," *IEEE Access*, Vol. 13, pp. 86799-86811, 2025. <https://doi.org/10.1109/ACCESS.2025.3570792>
- [9] A. Dey, A. Anand, S. Samanta, B. K. Sah, and S. Biswas, "Attention-Based AdaptSepCX Network for Effective Student Action Recognition in Online Learning," *Procedia Computer Science*, Vol. 233, pp. 164-174, 2024. <https://doi.org/10.1016/j.procs.2024.03.206>
- [10] A. Apicella, P. Arpaia, M. Frosolone, G. Improta, N. Moccaldi, and A. Pollastro, "EEG-Based Measurement System for Monitoring Student Engagement in Learning 4.0," *Scientific Reports*, Vol. 12, 5857, 2022. <https://doi.org/10.1038/s41598-022-09578-y>
- [11] L.-M. Vortmann, S. Ceh, and F. Putze, "Multimodal EEG and Eye Tracking Feature Fusion Approaches for Attention Classification in Hybrid BCIs," *Frontiers in Computer Science*, Vol. 4, 780580, 2022. <https://doi.org/10.3389/fcomp.2022.780580>
- [12] Y. Li, Y. Li, X. He, J. Fang, C. Zhou, and C. Liu, "Learner's Cognitive State Recognition Based on Multimodal Physiological Signal Fusion," *Applied Intelligence*, Vol. 55, 127, 2025. <https://doi.org/10.1007/s10489-024-05958-1>
- [13] S.-Y. Kim, S.-J. Urm, S.-Y. Yoo, S.-J. Kim, and K.-M. Lee, "Application of Sign Language Gesture Recognition Using Mediapipe and LSTM," *Journal of Digital Contents Society*, Vol. 24, No. 1, pp. 111-119, 2023. <https://doi.org/10.9728/dcs.2023.24.1.111>
- [14] M. Li, L. Tang, L. Ma, H. Zhao, J. Hu, and Y. Wei, "An Analysis Model of Learners' Online Learning Status Based on Deep Neural Network and Multi-Dimensional Information Fusion," *Computer Modeling in Engineering and Sciences*, Vol. 135, No. 3, pp. 2349-2371, 2022. <https://doi.org/10.32604/cmescs.2023.022604>



유민경 (Min-Kyeong Yu)

2021년~현 재: 덕성여자대학교 컴퓨터공학전공 학사과정
※ 관심분야: 딥러닝, 컴퓨터비전, 백엔드 시스템 엔지니어링



이하늘 (Ha-Neul Lee)

2021년~현 재: 덕성여자대학교 컴퓨터공학전공 학사과정
※ 관심분야: 백엔드, 인공지능, 컴퓨터비전



오지연 (Ji-Yeon Oh)

2021년~현 재: 덕성여자대학교 컴퓨터공학전공 학사과정
※ 관심분야: 백엔드, 인공지능, 컴퓨터비전



박유정 (Yoo-Jung Park)

2021년~현 재: 덕성여자대학교 컴퓨터공학전공 학사과정
※ 관심분야: 웹앱, UX/UI, 인공지능



양규리 (Gyu-Ri Yang)

2021년~현 재: 덕성여자대학교 컴퓨터공학전공 학사과정
※ 관심분야: 딥러닝, 안드로이드, 클라우드 컴퓨팅



이경미(Kyoung-Mi Lee)

1996년 : 연세대학교 대학원 (이학석사)
2001년 : Univ. of Iowa, Dept. of
Computer Science (전산학박
사-지능형멀티미디어)

2003년~현 재: 덕성여자대학교 컴퓨터공학전공 교수

※ 관심분야 : 영상처리, 컴퓨터비전, 패턴인식